

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-268012

(43)Date of publication of application : 29.09.2000

(51)Int.Cl.

G06F 15/177

G06F 15/16

(21)Application number : 11-067554

(71)Applicant : NEG CORP

(22)Date of filing : 12.03.1999

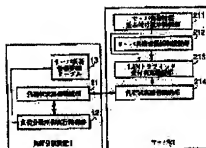
(72)Inventor : HIWATARI TERUMI

(54) METHOD AND DEVICE FOR DISTRIBUTING LOAD IN CLIENT SERVER SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a sure high-reliability service to a client by enabling operation with an optimum load.

SOLUTION: Concerning a method for distributing a server load, in a client server system, each server 21 weights its own load condition to a CPU load, a job priority, a number of execution jobs and a number of job input queues respectively, allocates a threshold value (server load information weighting setting function part 211) and periodically compares this threshold value with sampled (server load information sampling function part 212) load information so as to perform load judgement for instructing the acceptance or interruption of a local area network(LAN) traffic (LAN traffic acceptance discriminating function part 213). By timely transmitting load control information generated according to the result of the above load judgement to a load distributing device 1 connected to a LAN, the transmission of the LAN traffic to the relevant server is controlled (load distribution control executing function part 12).



LEGAL STATUS

[Date of request for examination] 01.03.2000

[Date of sending the examiner's decision of rejection] 02.04.2002

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-268012

(P2000-268012A)

(43) 公開日 平成12年9月29日(2000.9.29)

(51) Int.Cl.	識別記号	F I	子コード(参考)
G 0 6 F 15/177	6 7 4	G 0 6 F 15/177	6 7 4 B 5 B 0 4 5
15/16	6 2 0	15/16	6 2 0 B

審査請求 有 請求項の数 4 O L (全 10 頁)

(21) 出願番号 特願平11-87554

(22) 出願日 平成11年3月12日(1999.3.12)

(71) 出願人 00004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 日渡 真美

東京都港区芝五丁目7番1号 日本電気株式会社内

(74) 代理人 100108578

弁理士 高橋 昭男 (外3名)

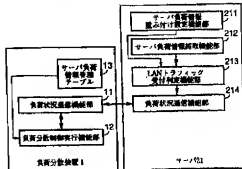
Fターム(参考) 5B045 AA03 B012 B028 B047 G004

(54) 【発明の名称】 クライアントサーバシステムにおけるサーバ負荷の分散方法ならびに装置

(57) 【要約】

【課題】 最適な負荷で運用でき、クライアントに対し確実で信頼性の高いサービスを提供する。

【解決手段】 クライアントサーバシステムにおいて、各サーバ21(22, 23)は自身の負荷状況をCPU負荷、ジョブ優先順位、実行ジョブ数、ジョブ入力待ち行列数のそれぞれに重み付けを行い、また、閾値を割当て(サーバ負荷情報重み付け設定機能部211)、この閾値と採取(サーバ負荷情報採取機能部212)された負荷情報を定期的に比較することにより、LANTラフィックの受け入れ、あるいは中断を指示する負荷判断を行う(LANTラフィック受付判定機能部213)。そして、LANT接続された負荷分散装置1に前記負荷判断の結果に従い生成される負荷制御情報をタイムリーに送信することにより、該当サーバへのLANTラフィックの送信制御を行う(負荷分散制御実行機能部12)。



【発明の要約】

【請求項1】 LAN接続環境下におけるクライアントサーバシステムにおいて、サーバ負荷の要素をアプリケーションに応じて定義される優先順位毎、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数に分解し、前記各サーバ毎、前記分解された要素のそれぞれにつき重み付けを行うと共に閾値を割り当て、この閾値と採取される負荷状況を定期的に比較することにより、LANTラフィックの受け入れ、拒否のための負荷判断を行い、その結果をLAN接続される負荷分散装置に通知することにより、前記重み付け設定された負荷要素に従い該当サーバもしくは他のサーバに対してLANTラフィックの送信制御を促すことを特徴とするサーバ負荷の分散方法。

【請求項2】 サーバ負荷の要素を、アプリケーションに応じて定義される優先順位毎、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数に分解し、前記それぞれの要素に基づきLANTラフィックの送信制御を行うクライアントサーバシステムにおいて、前記分解された要素のそれぞれにつき重み付けを行うと共に閾値を割り当て、この閾値と採取される負荷情報とを定期的に比較することにより、LANTラフィックの受け入れ、拒否のための負荷判断を行い、LAN接続される負荷分散装置に通知するサーバと、前記サーバによる負荷判断によっては内蔵するサーバ負荷管理テーブルを更新し前記重み付け設定された負荷要素に従い該当サーバもしくは他のサーバに対し、LANTラフィックの送信制御を行う負荷分散装置とを備えることを特徴とするサーバ負荷の分散装置。

【請求項3】 前記サーバ負荷管理テーブルは、管理すべきそれぞれのサーバで使用される優先度が付されたアプリケーション毎、CPU負荷ステータス、実行ジョブ負荷ステータス、入力ジョブ待ち行列負荷ステータスの各情報設定領域から成り、前記負荷分散装置により参照され、前記サーバ中の負荷情報重み付け設定手段で設定された重みに従い順序でLANTラフィックの受け入れ判断を行うことを特徴とする請求項2記載のサーバ負荷の分散装置。

【請求項4】 LAN接続環境下におけるクライアントサーバシステムにおいて、サーバ負荷の要素をアプリケーションに応じて定義される優先順位毎、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数に分解し、前記分解された要素のそれぞれにつき重み付けを行うと共に閾値を割り当て前記サーバ中のサーバ負荷情報重み付け設定手段と、前記CPU使用率、実行ジョブ数、ジョブ入力待ち行列数から成るサーバ負荷状況を定期的に採取する前記サーバ中のサーバ負荷情報採取手段と、前記サーバ負荷情報重み付け設定手段により定義された閾値と前記サーバ負荷情報採取手段により採取されたサーバ負荷状況を定期的に比較し、LANTラフィックの受け入

れ、もしくは拒否する前記サーバ中のLANTラフィック受付判定手段と、前記比較の結果によってはLANTラフィック中断のための負荷制御情報もしくはLANTラフィック受付再開のための負荷制御情報を、LAN接続される負荷制御装置に伝える前記サーバ中の負荷状況通信手段と、前記各負荷制御情報を受信して負荷分散制御実行手段に伝える前記負荷制御装置中の負荷状況通信手段と、LANTラフィック受付中断のための負荷制御情報を発信した場合、自身で管理しているサーバ負荷管理情報テーブルにおける該当サーバの関連ステータスを送信不可能状態に設定し、他のサーバでクライアントからのジョブを受け付け可能なサーバを探してルーティングを行い、LANTラフィック受付再開のための負荷制御情報を発信した場合、自身で管理しているサーバ負荷管理情報テーブルにおける該当サーバの関連ステータスを送信可能な状態に設定し、該当サーバに対するLANTラフィックの送信制御を促す前記負荷分散装置中の負荷分散制御実行手段とを備えることを特徴とするサーバ負荷の分散装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、クライアントサーバシステムにおけるサーバ負荷の分散方法ならびに装置に関する。

【0002】

【従来の技術】コンピュータをネットワーク化し、ネットワーク全体でコンピュータリソースを確保させるのがネットワークコンピュータである。ネットワークコンピュータにおいて、ネットワークを構成する各コンピュータは業務を分散化し、かつ相互に利用しあう分散処理が可能になる。サービスを提供するサーバや、サーバの機能を利用するクライアントがネットワークコンピュータ構成要素になるため、このネットワークコンピュータのことをクライアントサーバコンピュータと呼ぶことがある。上述したクライアントサーバコンピュータを実現するクライアントサーバシステムは、要求に応じて所定の処理を行うサーバコンピュータと、そのサーバコンピュータに対して処理を要求する複数のクライアントコンピュータと、これらクライアント、サーバコンピュータ間を相互に接続するLAN(Local Area Network)等通信回線から成る。

【0003】

【発明が解決しようとする課題】ところで、上述したクライアントサーバシステムにおいて、1台のサーバコンピュータの処理能力には限界があるため、サーバコンピュータを複数台で構成することが多く、この場合、特定のサーバに要求が集中しないように負荷分散のための手段が必要となる。通常はCPU負荷によってのみ負荷状況を判断し、負荷分散をはかっていた。また負荷分散装

置を接続し、サーバ間におけるレスポンスタイムを計測し、そのタイムに依り負荷状況を判断するシステムもあった。そのため、サーバ負荷は、CPU使用率、あるいはレスポンスタイムによってのみサーバの負荷が決定される適当なサーバへの割り当てられるため、サービスの重要度、種類などの負荷状況が的確に把握できず、従ってそのサーバの処理がゆがみずとも信頼性の高いサービスを提供するものではなかった。この発明は上記事情に鑑みてなされたものであり、サーバ負荷の要素を使用アプリケーションに応じて定義される優先順位、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数に分解し、それぞれについてLANTラフィックの流量を制御することにより、最適な負荷で運用でき、クライアントに対し確実に信頼性の高いサービスを提供できる、クライアントサーバシステムにおけるサーバ負荷の分散方法ならびに装置を提供することを目的とする。

【0004】

【課題を解決するための手段】上述した課題を解決するために請求項1記載のサーバ負荷の分散方法は、LAN接続環境下におけるクライアントサーバシステムにおいて、サーバ負荷の要素をアプリケーションに応じて定義される優先順位、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数に分解し、前記各サーバは、前記分解された要素のそれぞれにつき重みづけを行うと共に閾値を割り当て、この閾値と採取される負荷状況を定期的に比較することにより、LANTラフィックの受け入れ、拒否のための負荷判断を行い、その結果をLAN接続される負荷分散装置に通知することにより、前記重みづけ設定された負荷要素に依り該当サーバもしくは他のサーバに対してLANTラフィックの送信制御を促すことを特徴とする。

【0005】請求項2記載のサーバ負荷の分散装置は、サーバ負荷の要素をアプリケーションに応じて定義される優先順位、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数に分解し、前記それぞれの要素に基づきLANTラフィックの送信制御を行うクライアントサーバシステムにおいて、前記分解された要素のそれぞれにつき重みづけを行うと共に閾値を割り当て、この閾値と採取される負荷情報とを定期的に比較することにより、LANTラフィックの受け入れ、拒否のための負荷判断を行い、LAN接続される負荷分散装置に通知するサーバと、前記サーバによる負荷判断に基づいて内蔵するサーバ負荷管理テーブルを更新し前記重みづけ設定された負荷要素に依り該当サーバもしくは他のサーバに対し、LANTラフィックの送信制御を行う負荷分散装置とを備えることを特徴とする。また、請求項3記載のサーバ負荷の分散装置は、前記サーバ負荷管理テーブルは、管理すべきそれぞれのサーバで使用された優先度が付されたアプリケーション毎、CPU負荷ステータス、実行ジョブ負荷ステータス、入力ジョブ待ち行列負荷ステータス

の各情報設定領域から成り、前記負荷分散装置により参照され、前記サーバ中の負荷情報重み付け設定手段で設定された重みに依り順序でLANTラフィックの受け入れ判断を行うことも特徴とする。

【0006】更に請求項4記載の負荷の分散装置は、LAN接続環境下におけるクライアントサーバシステムにおいて、サーバ負荷の要素をアプリケーションに応じて定義される優先順位、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数に分解し、前記サーバ中のサーバ負荷情報重み付け設定手段と、前記CPU使用率、実行ジョブ数、ジョブ入力待ち行列数から成るサーバ負荷状況を定期的に採取する前記サーバ中のサーバ負荷情報採取手段と、前記サーバ負荷情報重み付け設定手段により定義された閾値と前記サーバ負荷情報採取手段により採取されたサーバ負荷状況を定期的に比較し、LANTラフィックの受け入れ、もしくは拒否する前記サーバ中のLANTラフィック受付判定手段と、前記比較の結果によってはLANTラフィック中断のための負荷制御情報もしくはLANTラフィック受け付けの再開のための負荷制御情報を、LAN接続される負荷制御装置に伝える前記サーバ中の負荷状況送信手段と、前記各負荷制御情報を受信して負荷分散制御実行手段に伝える前記負荷制御装置中の負荷状況通信手段と、LANラフィック受け付け中断のための負荷制御情報を受信した場合、自身で管理しているサーバ負荷管理テーブルにおける該当サーバの関連するステータスを送信可能な状態に設定し、該当サーバに対するLANTラフィックの送信制御を促す前記負荷分散装置中の負荷分散制御実行手段とを具備することを特徴とする。

【0007】上述した構成において、この発明は、サーバ群へのLANTラフィックの負荷分散をLAN環境で行う場合に、サーバ群のLANTラフィック負荷分散をサーバ群と負荷分散装置が協調して行うことを特徴とする。各サーバは自身の負荷状況をCPU負荷、ジョブ優先順位、実行ジョブ数、ジョブ入力待ち行列数のそれぞれに重みづけを行い、また、閾値を割り当て、この閾値と採取された負荷情報を定期的に比較することにより、LANTラフィックの受け入れ、あるいは中断を指示する負荷判断を行う。そして、LAN接続された負荷分散装置に前記負荷判断の結果に依り生成される負荷制御情報をタイムリーに送信することにより、該当サーバへのLANTラフィックの送信制御を行う。このことにより、サーバ群のLANTラフィック負荷分散を迅速かつ効率よく行うことを可能とする。本発明により、サー

パ群は最悪な負荷で運用でき、クライアントに対し、確実に信頼性の高いサービスを提供できる。

【0008】

【発明の実施の形態】図1は本発明が採用されるクライアントサーバシステムの構成を示すブロック図である。図1において、クライアントサーバシステムは、負荷分散装置1とサーバ群2、及び図示せぬクライアントコンピュータがLAN回線3經由で接続される。本発明の負荷分散方法が具現化されるサーバ群2はサーバ21、22、25から成る。負荷分散装置1は、LANTラフィッ 10 ックをサーバ群2を構成する各サーバ21、22、23の負荷状況を見ながら分散制御する。

【0009】図2は本発明のクライアントサーバシステムにおける負荷の分散装置の実施形態を示すブロック図である。図において、図1に示す番号と重複するブロックは図1のそれと同じとする。図において、サーバ群2の一つであるサーバ21は、サーバ負荷情報重み付け設定機能部211、サーバ負荷情報採取機能部212、LANTラフィッ 20 ック受付判定機能部213と負荷状況通信機能部214から構成される。サーバ22、25もサーバ21と同様の構成を持つ。本発明の特徴は、サーバ負荷の要素をアプリケーションに応じて定義される優先順位、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数として分解し、それぞれについてLANTラフィックの流量を制御を行うことにある。サーバ負荷情報重み付け設定機能部211は、分解された要素のそれぞれにつき後述する重み付けを行うと共に後述する閾値を割り当て、LANTラフィッ 30 ック受付判定機能部213に供給する。また、サーバ負荷情報採取機能部212は、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数から成るサーバ負荷状況を定期的に採取してLANTラフィック受付判定機能部213に供給する。LANTラフィック受付判定機能部213は、サーバ負荷情報重み付け設定機能部211により定義された閾値とサーバ負荷情報採取機能部212により採取されたサーバ負荷状況を定期的に比較し、LANTラフィッ 40 ックの受け入れ、もしくは拒否し、その負荷制御情報として負荷状況通信機能部214に供給する。負荷状況通信機能部214は、LANTラフィック受付判定機能部213による比較操作の結果によってはLANTラフィック制御のための負荷制御情報もしくはLANTラフィック受付再開のための負荷制御情報を、LAN接続される負荷分散装置1に伝送する。

【0010】負荷分散装置1は、負荷状況通信機能部11と、負荷分散制御実行機能部12と、サーバ負荷管理テーブル13で構成される。負荷状況通信機能部11は、サーバ21の負荷状況通信機能部214から送信される負荷制御情報を受信して負荷分散制御実行機能部12に伝送する。負荷分散制御実行機能部12は、LANTラフィック受付中 50 断のための負荷制御情報を受信した場合、自身で管理しているサーバ負荷管理情報テーブルにおける該当サーバ

の関連ステータスを送信不可状態に設定し、他のサーバでクライアントからのジョブを受け付け可能なサーバを探してルーティングを行い、LANTラフィック受付再開のための負荷制御情報を受信した場合、自身で管理しているサーバ負荷管理情報テーブル13における該当サーバの関連するステータスを送信可能な状態に設定し、該当サーバに対するLANTラフィックの送信制御を促す。サーバ負荷管理情報テーブル13は、図5にその一例を示すように、管理すべきそれぞれのサーバで採用される優先度が付されたアプリケーション毎、CPU負荷ステータス、実行ジョブ負荷ステータス、入力ジョブ待ち行列負荷ステータスの各情報設定領域から成り、負荷分散装置1の負荷分散制御実行機能部12により参照され、サーバ21中の負荷情報重み付け設定機能部211で設定された重み付けに従って順次でLANTラフィックの受け入れ判断が行われる。

【0011】図3乃至図5は本発明の実施形態の動作を説明するために引用した図であり、それぞれ、サーバ負荷情報重み付け設定機能部211により設定される重み付けの一例を表形式で示す表1、サーバ負荷情報採取機能部212により採取される負荷採取状況の一例を表形式で示す表2、サーバ負荷情報管理テーブル13のデータ構造を表形式で示す表3である。図3に示す表1において、重み付け情報は、サーバ負荷情報領域に、優先順位、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数から成る。これら各要素は、CPU使用率、実行ジョブ数およびジョブ入力待ち行列数の順にLANTラフィック受け入れのために負荷情報の重み付けがなされ、また、各要素に対し、それぞれ上限、下限閾値が設定される。更に、ジョブ受付限界値およびジョブ受付再開閾値も設定されている。尚、使用されるアプリケーション毎、例えば基幹業務は優先度1、WWW(World Wide Web)アクセスは優先度10に設定されるものとする。図4に示す表2は、サーバ負荷情報採取機能部212がある一定間隔でジョブ優先順位毎にCPU使用率、実行ジョブ数、入力ジョブ待ち行列数を採取したものである。図5に示す表3は、負荷分散装置1の負荷分散制御実行機能部12が管理しているサーバ負荷管理情報テーブル13であり、管理すべきサーバ毎のCPU負荷、実行ジョブ負荷および入力ジョブ待ち行列負荷のそれぞれについて、LANTラフィック受け入れ可否の状態を表している。LANTラフィックの受け入れ判断は負荷情報重み付け設定機能部211で設定される重み付け情報に従い、CPU負荷、実行ジョブ負荷、入力ジョブ待ち行列負荷の順に行われる。図6、図7は本発明の動作を説明するために引用したフローチャートであり、サーバ21の動作手順、負荷分散装置1の負荷分散制御実行機能部12の動作手順をそれぞれフローチャートで示した図である。

【0012】以下、図3乃至図7を参照しながら図1、

図2に示す本発明実施形態の動作について詳細に説明する。本発明の特徴は、サーバ負荷の要素をアプリケーションに応じて定義される優先順位、CPU使用率、実行ジョブ数、ジョブ入力待ち行列数として分解し、それぞれについてLANTラフィックの流量を制御を行うことにあることは上述したとおりである。そこで、サーバ負荷情報重み付け機能設定部211は、サーバ負荷状況を各要素毎それぞれに重み付けし(ステップS61)、また、閾値として設定定義する(ステップS62)。ここでは図3に表2>として示すように、各要素毎の重み付けに關し、CPU使用率、実行ジョブ数、入力ジョブ待ち行列数の順に重み付けし、また、優先度1が付与されたアプリケーションに關し、CPU使用率の上限閾値を30%、下限閾値を15%、実行ジョブ数の上限閾値を60%、下限閾値を40%、入力ジョブ待ち行列の上限閾値を20%、下限閾値を10%とし、また、優先度2が付与されたアプリケーションに關し、CPU使用率の上限閾値を20%、下限閾値を7%、実行ジョブ数の上限閾値を40%、下限閾値を20%、入力ジョブ待ち行列の上限閾値を15%、下限閾値を10%とし、更に、優先度10が付与されたアプリケーションに關し、CPU使用率の上限閾値を5%、下限閾値を5%、実行ジョブ数の上限閾値を50%、下限閾値を40%、入力ジョブ待ち行列の上限閾値を80%、下限閾値を50%とし、後述するLANTラフィック受付判定機能部213における受付判定処理のため各要素毎の受付限界閾値、あるいは再開閾値としている。【0013】一方、サーバ負荷情報採取機能部212は、サーバ負荷状況をCPU負荷、ジョブ優先順位別の実行ジョブ数およびジョブ入力待ち行列数を定期的に採得し(ステップS63)、LANTラフィック受付判定機能部213に供給する。採取されたサーバ負荷状況は、ここでは図4に表2>として示す。ここでは、優先度1のジョブに關し、CPU使用率2.4%、実行ジョブ数43、入力ジョブ待ち行列数9とし、また、優先度2のジョブに關し、CPU使用率21%、実行ジョブ数38、入力ジョブ待ち行列数5とし、また、優先度10のジョブに關し、CPU使用率2.4%、実行ジョブ数43、入力ジョブ待ち行列数9とする。LANTラフィック受付判定機能部213は、サーバ負荷情報重み付け機能部211によって定義された各要素毎の受付限界閾値、受付再開閾値と、サーバ負荷採取機能部212により採取された負荷状況を定期的に比較し、LANTラフィックの受付/再開、あるいは拒否を判定する(ステップS64)。LANTラフィック受付判定機能部213により、サーバ負荷情報重み付け機能部211で定義された受付限界閾値を越えたことが確認された場合、負荷状況通信機能部214は、負荷分散装置1に対してLANTラフィック受付中断のための負荷制御情報を送信し(ステップS65)、また、受付再開閾値に負荷が下がった場合、LANTラフィックの受付再開のための負荷制御情報を送信

する(ステップS66)。定義された閾値の範囲にある場合は該当サーバでのLANTラフィックの受け入れを許可する(ステップS67)。

【0014】負荷分散装置1は、負荷状況通信機能部11でサーバ21の負荷状況通信機能部214から負荷制御情報を受信すると、その情報を負荷分散制御実行機能部12に転送する。また負荷分散制御実行機能部12により生成された後述する負荷分散制御情報を指示されたサーバ21の負荷状況通信機能部214に送信する。負荷分散制御実行機能部12は、サーバ21から送信される負荷制御情報をチェックし(ステップS71)、ここでLANTラフィック受付中断の負荷制御情報を受信した場合、負荷分散装置1内で管理している該当サーバ3の関連するステータスを送信不可状態に変更し(ステップS72)、受付確認メッセージをサーバ21に返答する(ステップS73)。更に、サーバ群2の他のサーバでクライアントからのジョブを受付可能なサーバをサーバ負荷情報管理テーブル13からラウンドロビン検索ルーティングする(ステップS74)。一方、サーバ21からLANTラフィック受付再開の負荷制御情報を受信すると、負荷分散装置1内で管理している該当サーバ214の関連するステータスを送信可能状態に変更し(ステップS75)、受付確認メッセージをサーバ3に返答する(ステップS76)。図5に示す表3>は、負荷分散装置1の負荷分散制御実行機能部12が管理しているサーバ負荷情報管理テーブル13であり、管理すべきサーバ毎のCPU負荷、実行ジョブ負荷および入力ジョブ待ち行列負荷のそれぞれについて、LANTラフィック受け入れ可否の状態が示されている。例えば、サーバ01の優先度1が割り振られたジョブはCPU負荷ステータスが受け入れ可、実行ジョブ負荷ステータスが受け入れ可、入力ジョブ待ち行列負荷ステータスが受け入れ可になっており、また、サーバ02の優先度8が割り振られたジョブはCPU負荷ステータス、実行ジョブ負荷ステータス、入力ジョブ待ち行列負荷ステータス共に受け入れ可の状態になっている。

【0015】以上説明のように本発明は、サーバ負荷の要素を使用アプリケーションに応じて定義される優先順位、CPU使用率、実行ジョブ数、ジョブ入力待ち行列に分解し、それぞれについてLANTラフィックの流量を制御するものであり、このことにより、最適な負荷で運用できる。

【0016】

【発明の効果】以上説明のように本発明によれば、サーバ毎の負荷分散の制御をサーバと負荷分散装置が協調してサービスの重要度、種類に応じて詳細に制御するため、サーバ群は最適な負荷で運用でき柔軟に信頼性の高いサービスを提供する。

【図面の簡単な説明】

【図1】 本発明が採用されるクライアントサーバシス

テムの構成例を示すブロック図である。

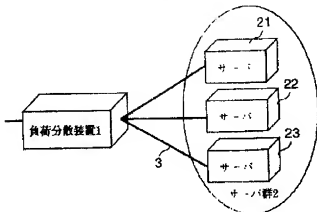
【図2】 本発明の実施形態を示すブロック図である。

【図3】 図2に示す重み付け設定機能部により定義される、重み付け、閾値設定例の一例を<表1>として示す図である。

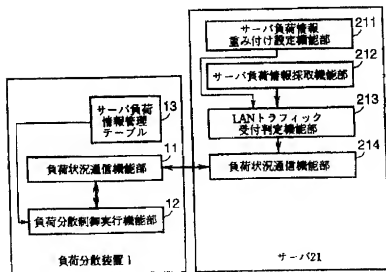
【図4】 図2に示す負荷情報採取機能部により採取される負荷情報の一例を<表2>として示す図である。

【図5】 図2に示す負荷分散装置により管理されるサーバ負荷情報管理テーブルのデータ構造の一例を<表3>として示す図である。

【図1】



【図2】



【図6】 図2におけるサーバの動作を説明するために引用したフローチャートである。

【図7】 図2における負荷分散制御実行機能部の動作を説明するために引用したフローチャートである。

【符号の説明】

1…負荷分散装置、2(21, 22, 23)…サーバ、11(214)
…負荷状況通信機能部、12…負荷分散制御実行機能部、
13…サーバ負荷情報管理テーブル、211…サーバ負荷情
報重み付け設定機能部、212…サーバ負荷情報採取機
部、213…LANトラフィック受付判定機能部、
214…サーバ負荷情報採取機能部

【 図3 】

< 表 1 > 重み付け・閾値設定

	CPU使用率 上限しきい値 下限しきい値	実行ジョブ数 上限しきい値 下限しきい値	入力ジョブ待ち行列数 上限しきい値 下限しきい値
優先度 1	3 0 % 1 5 %	6 0 4 0	2 0 1 0
優先度 2	2 0 % 7 %	4 0 2 0	1 5 7
優先度 10	5 % 4 %	5 0 4 0	8 0 5 0

【 図4 】

< 表 2 > 負荷採取状況

	CPU使用率	実行ジョブ数	入力ジョブ待ち行列数
優先度 1	2 4 %	4 3	9
優先度 2	2 1 %	3 6	5
⋮	⋮	⋮	⋮
優先度 10	7 %	5 7	3 5

【 図5 】

<表3> サーバ負荷情報管理テーブル13

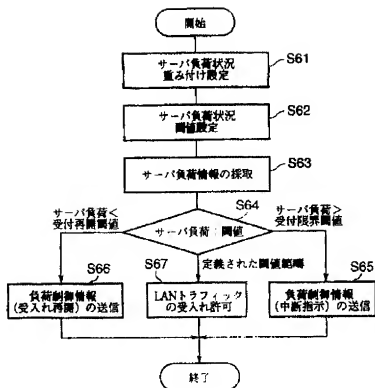
サーバ名	優先度	CPU負荷 ステイタス	実行ジョブ 負荷ステイタス	入力ジョブ待ち行列 負荷ステイタス
SVR01	1	受入可	受入不可	受入可
	2	受入不可	受入可	受入可

	10	受入可	受入可	受入不可
SVR02	1	受入不可	受入可	受入可

	8	受入可	受入可	受入可
SVR03	1	受入可	受入可	受入可

	10	受入不可	受入不可	受入不可

【 図6 】



【 図7 】

